Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania* (*Mundinia*) *enriettii*, Isolate CUR178, Strain LV763

Hatim Almutairi,[a,b] Michael D. Urbaniak,[a] Michelle D. Bates,[a] Vanete Thomaz-Soccol,[c] Waleed S. Al-Salem,[b] Rod J. Dillon,[a] Paul A. Bates,[a] ⓘDerek Gatherer[a]

[a]Division of Biomedical & Life Sciences, Faculty of Health & Medicine, Lancaster University, Lancaster, United Kingdom
[b]Ministry of Health, Riyadh, Saudi Arabia
[c]Laboratório de Biologia Molecular, Programa de Pós Graduação em Engenharia de Bioprocessos e Biotecnologia, Universidade Federal do Paraná, Curitiba, Brazil

ABSTRACT *Leishmania* (*Mundinia*) *enriettii* is a parasitic kinetoplastid first isolated from a guinea pig in Brazil in 1946. We present the complete genome sequence of *L.* (*M.*) *enriettii*, isolate CUR178, strain LV763, sequenced using combined short-read and long-read technologies. This will facilitate a greater understanding of the genome diversity within *L.* (*M.*) *enriettii*.

*L*eishmania enriettii was first isolated from a guinea pig (*Cavia porcellus*) in Brazil in 1946 and formally described 2 years later (1). The group of related species previously known as the *L. enriettii* complex is now referred to as the subgenus *Mundinia* (2, 3). A previous genome sequence of *L.* (*M.*) *enriettii*, isolate LEM3045 (MCAV/BR/1995/CUR3), was assembled but with a large number of unplaced contigs, higher gap content, and relatively lower N_{50} value compared to other *Leishmania* genomes (2). Additional *L.* (*M.*) *enriettii* isolates have been associated with leishmaniasis lesions in guinea pigs in the Curitiba Metropolitan Region of southern Brazil, 50 years after the first case (4, 5). We report here the complete genome assembly and annotation of *L.* (*M.*) *enriettii*, isolate CUR178, strain LV763 (WHO code MCAV/BR/2001/CUR178; LV763). This will contribute to our understanding of the biology of *L.* (*M.*) *enriettii* and the genomic diversity existing in this species.

We obtained intracellular amastigote parasites from six naturally infected guinea pigs from the rural area of Mandirituba (state of Paraná, Brazil). The parasites were grown using an *in vitro* culture system previously developed for *Leishmania* (*Mundinia*) *orientalis* axenic amastigotes (6) in Schneider's insect medium at 26°C as promastigotes, then in M199 medium supplemented with 10% fetal calf serum (FCS), 2% stable human urine, 1% basal medium Eagle vitamins, and 25 μg/ml gentamicin sulfate, with subpassage to fresh medium every 4 days to sustain the parasite growth and viability. DNA was extracted and purified using a Qiagen DNeasy blood and tissue kit using the spin column protocol, according to the manufacturer's instructions. The extracted DNA concentration was assessed using a Qubit fluorometer, microplate reader, and agarose gel electrophoresis. All sequencing libraries were based on the same extracted DNA sample to avoid any inconsistency.

Short-read library construction and sequencing were contracted to (i) BGI (Shenzhen, China) for DNBSEQ libraries, producing paired-end reads (270 bp and 500 bp) using the Illumina HiSeq platform, and (ii) Aberystwyth University (Aberystwyth, UK) for TruSeq Nano DNA libraries, producing paired-end reads (300 bp) using the Illumina MiSeq platform. We performed long-read library preparation and sequencing according to the Nanopore protocol (SQK-LSK109) on R9 flow cells (FLO-MIN106). The read quality was assessed using MultiQC (7).

We assembled the long reads using Flye (8), with default parameters, to generate chromosome-scale scaffolds. Then, using Minimap2 (9) and SAMtools (10), we mapped the short reads onto the assembled scaffolds to correct erroneous bases within long reads and create consensus sequences. After polishing the assembly with Pilon (11), another round of

A. Number of chromosomes and unplaced contigs

B. N50 and length of longest contig/scaffold

C. Genome length

- ■ *L. (M.) enriettii* LV763 (GCA_017916305.1- this paper)
- ■ *L. (M.) enriettii* LEM304 (GCA_000410755.2)
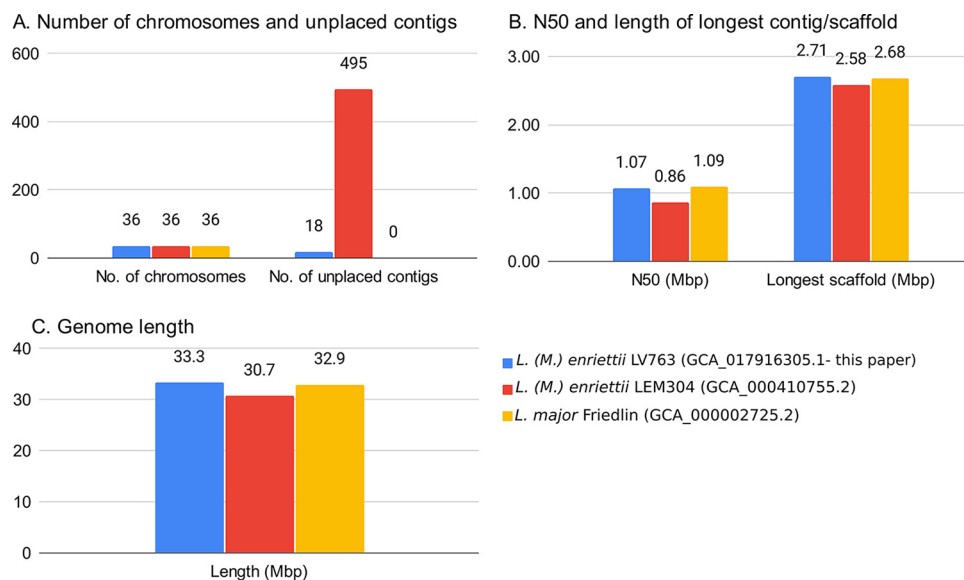- ■ *L. major* Friedlin (GCA_000002725.2)

FIG 1 Assembly comparison of *L. (M.) enriettii* LV763 with *L. (M.) enriettii* LEM3045 and *L. major* Friedlin.

consensus short read mapping was performed. Then, we removed the duplicated contigs and sorted the remainder according to length using Funannotate (12, 13). Finally, we separated the chimeric sequences and performed scaffolding using RaGOO (14) with the *Leishmania major* strain Friedlin genome (GenBank accession number GCA_000002725.2) (15) as a reference guide, aligning all 36 chromosomes for our assembly, thereby also determining the chromosome ends to be complete, with the exception of 18 unplaced contigs totaling 76,607 bp.

The analysis workflow for assembly, repeat masking, and annotation was performed using Snakemake (16); it is available online for reproducibility purposes (https://github.com/hatimalmutairi/LGAAP), including the software versions and parameters used (17). Figure 1 compares our assembly with other complete genomes.

We assessed the assembly completeness using BUSCO (18), with the lineage data set for the phylum *Euglenozoa*, containing 130 single-copy orthologs from 31 species, and we found that 123 of these orthologs were present (94.6% completeness). We carried out functional annotation and prediction using the MAKER2 (19) annotation pipeline in combination with

TABLE 1 Detailed summary metrics of the genome sequencing, assembly, and annotation for *L. (M.) enriettii* LV763

Feature(s)	Metric(s)
Total no. of reads	26,789,424
No. of MiSeq reads	5,060,124
No. of HiSeq reads	20,936,270
No. of MinION reads (read N_{50} [bp])	793,030 (12,070)
No. of bases (Gb)	19.41
Genome coverage ($\times$)	271.8
Total no. of scaffolds	54
Genome size (bp)	33,318,864
N_{50} (bp)	1,075,649
GC content (%)	59.60
No. of Ns (% of genome)	380 (0.001)
No. of genes	8,353
Gene density (no. of genes/Mb)	250.7
No. of exons	8,584
Mean gene length (bp)	1,897
Total length of CDSs[a] (Mb) (% of genome)	15.46 (46.40)

[a] CDSs, coding DNA sequences.

AUGUSTUS (20) gene prediction software, with the predictor trained on *Leishmania tarentolae*. Table 1 shows additional summary metrics for the sequencing, assembly, and annotation.

Data availability. The assembly and annotations are available under GenBank assembly accession number GCA_017916305.1. The master record for the whole-genome sequencing project is available under accession number JAFHKP000000000.1. The raw sequence reads are available under accession number PRJNA691534.

ACKNOWLEDGMENT

REFERENCES

1. Paraense WL. 1953. The spread of *Leishmania enriettii* through the body of the guineapig. Trans R Soc Trop Med Hyg 47:556–560. https://doi.org/10.1016/s0035-9203(53)80008-8.
2. Butenko A, Kostygov AY, Sádlová J, Kleschenko Y, Bečvář T, Podešvová L, Macedo DH, Žihala D, Lukeš J, Bates PA, Volf P, Opperdoes FR, Yurchenko V. 2019. Comparative genomics of *Leishmania* (*Mundinia*). BMC Genomics 20:726. https://doi.org/10.1186/s12864-019-6126-y.
3. Sereno D. 2019. *Leishmania* (*Mundinia*) spp.: from description to emergence as new human and animal *Leishmania* pathogens. New Microbes New Infect 30:100540. https://doi.org/10.1016/j.nmni.2019.100540.
4. Machado MI, Milder RV, Pacheco RS, Silva M, Braga RR, Lainson R. 1994. Naturally acquired infections with *Leishmania enriettii* Muniz and Medina 1948 in guinea-pigs from Sao Paulo, Brazil. Parasitology 109:135–138. https://doi.org/10.1017/S0031182000076241.
5. Thomaz-Soccol V, Pratlong F, Langue R, Castro E, Luz E, Dedet JP. 1996. New isolation of *Leishmania enriettii* Muniz and Medina, 1948 in Parana State, Brazil, 50 years after the first description, and isoenzymatic polymorphism of the *L. enriettii* taxon. Ann Trop Med Parasitol 90:491–495. https://doi.org/10.1080/00034983.1996.11813074.
6. Chanmol W, Jariyapan N, Somboon P, Bates MD, Bates PA. 2019. Axenic amastigote cultivation and in vitro development of *Leishmania orientalis*. Parasitol Res 118:1885–1897. https://doi.org/10.1007/s00436-019-06311-z.
7. Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354.
8. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
9. Li H. 2016. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
11. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
12. Li W-C, Wang T-F. 2021. PacBio long-read sequencing, assembly, and Funannotate reannotation of the complete genome of *Trichoderma reesei* QM6a. Methods Mol Biol 2234:311–329. https://doi.org/10.1007/978-1-0716-1048-0_21.
13. Palmer JM, Stajich JE. 2020. Funannotate v1.8.1: eukaryotic genome annotation. https://doi.org/10.5281/zenodo.1134477.
14. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 20:224. https://doi.org/10.1186/s13059-019-1829-6.
15. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream M-A, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RMR, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, et al. 2005. The genome of the kinetoplastid parasite, *Leishmania major*. Science 309:436–442. https://doi.org/10.1126/science.1112680.
16. Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, Wilm A, Holtgrewe M, Rahmann S, Nahnsen S, Köster J. 2021. Sustainable data analysis with Snakemake. F1000Res 10:33. https://doi.org/10.12688/f1000research.29032.2.
17. Almutairi H, Urbaniak MD, Bates MD, Jariyapan N, Kwakye-Nuako G, Thomaz-Soccol V, Al-Salem WS, Dillon RJ, Bates PA, Gatherer D. 2021. LGAAP: *Leishmaniinae* genome assembly and annotation pipeline. Microbiol Resour Announc 10:e00439-21. https://doi.org/10.1128/MRA.00439-21.
18. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
19. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491. https://doi.org/10.1186/1471-2105-12-491.
20. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.